Dear William Colgan,

Thank you for your constructive comments on our manuscript. We have addressed all the points you raised, and provide the detailed answers below. All technical corrections have been accepted as suggested, and are not listed in detail here.

Line / item / section	Comment	Answer
Authorship	I am surprised by such a concise author list for a data paper of this magnitude. While I appreciate that the authors describe assembling and digitizing existing online resources, the net effect of compiling such a data paper can result in citation siphoning from the underlying papers. Simply put, the tendency to cite this data paper will likely reduce the citations, and thus the visibility, of original works. This can be especially detrimental for early-career researchers and/or less well-known research groups. (e.g. https://www.nature.com/nature-index/news/review-articles-cause-dramatic-loss-in-citations-for-original-research) When we assembled the Løkkegaard et al. (2023) review paper and associated Mankoff et al. (2022) database, we made a deliberate effort to avoid citation siphoning by following CREDIT principles (Brand et al., 2015; https://doi.org/10.1087/20150211) and introducing explicit guidance in the Data Statement to ensure contributor credit. I gently point out that Jacquemart and Welty (2024) contribute to citation siphoning by simply citing the "Mankoff (2022)" Github working directory, rather than the recommended "Mankoff et al., 2022" GEUS Dataverse data citation that includes the >20 data authors (see: https://doi.org/10.22008/FK2/3BVF9V). The current authorship structure of both the paper and the Zenodo repository (both cited as "Jacquemart and Welty (2024)") makes no attempt to combat citation siphoning. Several times, the authors acknowledge the laborious nature of collecting ice temperature profiles as well as the hopes that glenglat will be adopted as a community database receiving new submissions, but the present authorship structure effectively asks researchers to provide their data without a clear path for receiving citation. Being a "data contributor" on Zenodo generates no formal citation or impact metric.	The short author list was by no means an attempt to draw citations away from anyone, but simply a result of many unanswered emails and calls for data sharing, which ultimately pushed us do the vast majority of the work by ourselves and by hand (i.e. M. Jacquemart & E. Welty). We have now ammended this by making all data curators and data collectors are now co-authors on the dataset. They were also able to suggest others that had made a meaningful contribution to the project, so the list has gotten a bit longer. We also invited everyone to become a co-author on the paper, but all dataset co-authors turned down that invitation. We convinced Guillem Carcanade and Marcus Gastaldello to join the author team for the publication because they contributed substantially larger efforts than others to the dataset (see author contributions). Beyond this, we have added a table with the names of all authors involved in the publications from which we sourced the data – as a reasonable but not perfect proxy of who might have been involved in collecting the original data – to the appendix. This is referenced in the introduction and in the paper acknowledgments. In the paper, we have outlined the following authorship policy: *Anyone who submits data to glenglat will be invited to become a co-author on future releases of the dataset (see detailed authorship policy at https://github.com/mjacqu/glenglat/tree/main?tab=readme-ov-file#authorship-policy).* To further highlight the efforts that led to the original data collection, we have also added columns borehole.investigators and borehole.funding, which can be used to store information about the people and institutions that carried out the original work as well as the funding bodies. These columns are intended rather for submitted datasets, and we did not go back to published sources to retrieve information to back-fill these columns. The co-authors on the dataset in its current form are: M. Jacquemart E. Welty G. Carcanade L. van Tricht G. Flowers S. Sugiyama T. Gurung R. Prinz J. Abermann J. Steiner

		M. Barandun O. Gagliardini C. Vincent L. Thompson Zang T. M. Gastaldello M. Hoelzle
	I gently point out that Jacquemart and Welty (2024) contribute to citation siphoning by simply citing the "Mankoff (2022)" Github working directory, rather than the recommended "Mankoff et al., 2022" GEUS Dataverse data citation that includes the >20 data authors (see: https://doi.org/10.22008/FK2/3BVF9V).	Added Mankoff et al. 2022 (https://doi.org/10.22008/FK2/3BVF9V) and Løkkegaard et al. 2023 (https://doi.org/10.5194/tc-17-3829-2023) as secondary sources, but kept the GitHub pages ("Mankoff 2022") as primary data source because the temperature profiles at Mankoff et al. 2022 are interpolated to 1 m depth, rather than the originally reported depths.
94	It seems a little inconsistent to "occasionally include" snow temperatures simply if they are alongside ice temperatures.	Yes, the wording to describe this was misleading. We intended to say that we did not specifically seek out shallow measurements, but we included them if they were included as part of a deeper firn or ice profile or presented immediately alongside firn or ice temperature. The columns borehole.ice_depth and the borehole.notes provide information to distinguish these. We have updated the statement in the paper to read as follows: *We disregarded publications that provided measurements that were made exclusively in seasonal snow, but did not remove shallow measurement points of deeper measurements in firn or ice. Where shallow measurements were provided alongside deeper ones, we did also include these.*
117	Would perhaps be helpful to have an overview figure of the glenglat architecture, data flow, through different software packages and file types.	We have added a figure that depicts the workflow as it is described in the text.
135	Is the metadata being described here the specific tags shown in Table 2? This can be made explicit.	The metadata we refer to here are described throughout tables 1-4. We have inserted the specific reference, and we referenced each table next to its description lower in the paragraph.
164	It would seem appropriate to have description of the digitization error uncertainty (i.e. Figure 10) here, as it is part of the Method and not Results/Discussion of the data base?	Thank you for this comment. The errors can only be estimated retrospectively, as a result of having the database, in our opinion, we do believe that they are therefore a result of sorts. We also prefer to discuss all the errors in one place, and we believe that the results/discussion section is more appropriate than the methods section.
205	Yes, it does seem relevant to provide in the database where boreholes are in the accumulation or ablation area.	We added a column borehole.mass_balance_area with values 'accumulation', 'ablation', 'equilibrium' and populated it from borehole.notes.
225	It would be useful to visualize/assess whether warm bias changes with time, as presumably more recent records have a greater likelihood of meltwater refreezing within firn, and releasing latent heat, than older records.	This is an excellent comment, and we hope that glenglat can serve to investigate exactly such questions. Work from places with long-term measurements does show that these trends are visible, but that also quantifying the exact changes is not trivial (see for example Gastaldello et al., 2024; EGUsphere). In that sense, we believe that such an analysis is beyond the scope of this paper. To highlight the issue and the consequences of englacial warming, however, we added a paragraph on the topic in the

		introduction. We have added the following: *Ongoing human-driven climate change is leading to substantial changes of englacial temperatures, evidence of which is clear in (few existing) repeat measurements. At Dôme du Gouter in the French Alps, for example, a warming of 1.5°C has been recorded at a depth of 50 m between 1994 and 2017 (Vincent et al., 2020). Long-term measurements at Golle Gnifetti (Swiss Alps) between 1991 and 2023 reveal the same amount of warming at a depth of 20 m (Gastaldello et al., 2024). This warming can have several consequences. For one, meltwater infiltrating cold firn can degrade or destroy the archive of past climatic conditions that can be stored in firn or ice cores. In steep terrain, warming at the glacier bed can lead to wide-spread sliding and destabilization of entire glaciers, increasing the probability of very large ice avalanches, similar to the one observed at Altels in Switzerland in 1895 (Heim, 1895; Faillettaz et al., 2011). Counterintuitively, warming can also lead some glaciers to cool, because their disappearing firn cover reduces the amount of englacial warming that stems from the latent heat release of refreezing meltwater (Gilbert et al., 2012, Huss & Fischer 2016, Irvine-Fynn et al., 2011). A similar effect may also occur in areas still covered by firn, if the refreezing meltwater creates impermeable ice layers that prevent water percolation into the firn, thereby (locally) limiting the latent heat release (Vincent et al., 2020).*
282	Perhaps describe a couple examples of prime targets for such a retroactive comparison?	Thanks for this suggestion. We added a few ideas (there are many options!). To identify possible locations, we filtered the database for measurements that have published uncertainties that are less than 0.1°C, are deep and have (had) ice temperatures that are predominantly cold, are not on known surge-type glaciers, and have not been measured since at least 20 years. To the text we added the following statement: *An identification procedure for high-value sites for repeat measurements could focus on sites with deep boreholes (e.g., more than 30 m), predominantly cold englacial temperatures, high measurement accuracies (low reported uncertainties), and a last measurement dating back twenty years or more. Ignoring surge-type glaciers, a search of glenglat reveals high-value repeat-measurement sites in many different regions and climates. Some examples that fall into this category include Illimani Volcano (Bolivia, 138 m, measured in 1999), Vavilov Glacier (Russian Arctic, up to 460 m, last measured in 1985), or Åsgårdfonna (Svalbard, 185 m, measured in 1993).*
315	Some citations for the cold-temperate transition being "often extracted" from radar data.	We have added references to this method from our existing sources (studies that compared borehole temperatures to radar data).
Table 3/4	Please clarify what "/**/" denotes in the file extension.	The ** indicates subdirectories that contain separate `profile.csv` and `measurement.csv` files from boreholes with many profiles (e.g., from automated loggers). The subdirectories are

		labeled source.id-glacier, where glacier is a simplified version of the glacier name (e.g. flowers2022-little-kluane). We have added this information to the relevant table captions.
Fig. 6	It may be helpful to indicate which profiles are temperate and cold-based in 6b, as temperate ice is an upper limit temperature. It might also be insightful to divide the comparison into 1950-1990 and post-1990 (or c. 2000 breakpoint) to highlight if the biases are different in the more recent period during which more meltwater is percolating into historical firn accumulation zones than in the past.	The figure is a comparison between (ERA5) surface temperature and borehole temperature at 15 m depth. The 15 m depth was chosen to avoid seasonal surface fluctuations while including as many shallower boreholes as possible. Since only 20% of boreholes are known to have reached the bed, segmenting by basal temperature would rule out too many boreholes. It also isn't clear to us how relevant this is to temperatures at 15 m depth. We tried segmenting by measurement year (>= 1990, > 1990) and the differences are not obvious enough to be conclusive. A great deal of careful statistical analysis would be needed to sort out whether the differences are climate-change-driven or a result of comparing completely different sets of borehole locations. We believe it is better to omit this from the figure than to suggest something that is unsubstantiated.
2.4	What do you do when the profile graphic to be digitized simply has a line graph, without specific points, and thus the individual measurements are not discretized?	Thank you for this comment. Indeed, the digitization (mainly the resulting depths) depends strongly on how the initial data was displayed. To clearly distinguish between line and point graphs, we have replaced 'digitized' in the profile.measurement_origin column with either 'digitized-discrete' and 'digitized-continuous'. As for how the digitization (and reproduction) error was calculated, we added the following to the text: *We calculated both these errors from profile pairs as the difference between their temperatures after interpolating (but never extrapolating) the temperatures of one to match the depths of the other.*
3.2	The comparison against ERA5 air temperature and precipitation is interesting, although I would caution that more caveats need to be provided with regards to interpreting the ERA5 data, which is known to be challenged in areas of complex topography -- where most glaciers are often found. For example, could ERA5 bias with elevation contribute to the apparent warm bias? Comparing modelled surface temperatures with observed ice temperatures is different from comparing observed surface temperatures with observed ice temperatures. Also, be explicit on the temporal subset of profiles in this comparison, which appears to be only post-1950 profiles.	This is valuable input, thank you. Indeed, the plot only includes profiles since 1960, which we have now specified in the caption *Englacial temperatures (1960 to present)*. We have also added additional information about the caveats of ERA-5 and other reanalysis products, now stating that: *The trends and boundaries in Fig. 7b need to be evaluated with respect to possible caveats of the chosen dataset, and are intended only to show broad patterns. ERA5-Land (and other reanalysis products) may not be able to represent the true variability of temperature fields or precipitation patterns in complex alpine terrain. The observed 'warm-bias' could therefore partly be due to biases within this dataset, and the 2-m air temperature is in reality not the true surface boundary condition. Beyond this, the chosen lapse-rate (in our case 6.5°C/km) changes the relationship between air temperature and englacial temperature, though we found that it mostly affects the positions of the data points relative to the 1:1 line, while the overall shape remains stable*

3.3	I am not sure this involved discussion of sampling bias adds much to the paper. Yes, a sampling bias clearly exists, it is not clear why it is important. Would the authors perhaps expand on how they envision -- at highest level -- the database potentially being used in such a way as these biases become important? For example, if used as a training dataset when simulating the global population of englacial ice temperatures, would the bias result in potentially over- or under-estimating either present day or historical mean ice temperatures?	We do believe it is important to lay out broadly what sampling biases might be present in the dataset. If we envision wanting to model the distribution or future evolution of englacial temperatures, knowledge of sampling biases is imporant. In other words, if we want to constrain cold-temperate boundary spatially, by elevation, by temperature or precipitation regime, or based on glacier dynamics, we need measurements that span the transition and range of conditions where glaciers exist. To clarify this, we have added the follwoing statement: *If we envision using global or regional models to constrain thermal regimes spatially, by elevation, by temperature or precipitation regime, or based on glacier dynamics, we ideally need measurements that span the transition and range of conditions where glaciers exist. In this context it is important to understand what data is present in a possible training dataset. In the following, we therefore briefly discuss the temporal and spatial patterns of the data in glenglat.*
3.4	There is a very nice discussion of the digitization error here. Perhaps readers would appreciate more discussion (or guidance) on the population of measurement errors reported in Figure 10, and whether some generalization can be made for an adopted value. Or if the mention of 0.14°C as the mean meant to reflect a suitable "characteristic value" for the measurement uncertainty population? It would also be helpful to have further discussion on how different types of errors interact or combine to potentially bias results, and how to mitigate this issue.	Thank you for this comment. We have re-written section 3.4 (Error analysis) and now provide a comprehensive discussion about the different sources of error, how these are reported in the literature, and how we can estimate their impact on the comparability and accuracy of the data contained in glenglat.
3.4	It would also be helpful for some discussion of depth uncertainty. The database is effectively T(z), and there is clearly uncertainty in z that is not captured at present. There is both digitization uncertainty and also measurement uncertainty, especially when boreholes do not reach the bed and have poorly constrained altitudes. Presumably, end users will need guidance on how to compare a historical depth temperature with modeled glacier geometry. For example, when writing Løkkegaard et al. (2023), we were urged to also express temperatures on fraction 0 to 1 depth scale that could be fit to modelled ice geometries.	See above, we have also included a paragraph discussing the depth uncertainty in the new section 3.4
Data license	I understand that the CC Attribution 4.0 International license ensures, among other things, "appropriate credit" to the current work (i.e. Jacquemart and Welty (2024)) and transparency for any changes to derivative products (https://creativecommons.org/licenses/by/4.0/deed.en). Authorship aside, I am not entirely sure if glenglat itself has met these data conditions with respect to the original works, particularly when digitization of a figure has created a derivative dataset of an original profile for which no DOI is available. Is it possible to make otherwise unavailable .pdf's available in a protected literature repository?	We would be delighted to make all the literature publicly available, but based on discussions with our librarians, this is not advised. We have added a statement in the code and data availability that we will gladly share any of the cited literature with anyone. There, and in the GitHub Readme, we have also added a statement that the license applies to data and software, but not to the data in the /sources folder: *Glenglat is licensed under Creative Commons Attribution 4.0 International, though the repository's license does not extend to figures, tables, maps, or text extracted from publications. These are included in the sources folder for*

	This is an approach I have seen in other community data compilations, which can also help elevate otherwise overlooked works. That could address the appropriate credit link to the original work.	*transparency and reproducibility. Full PDFs of the original sources will gladly be shared upon request at any time.*
	For transparency on the derivative product, perhaps showing the original figures as well as the original final with the digitized profile overlaid? This would highlight how well the dertivative fits the original.	The /sources folder on the GitHub repository contains the original figures. For data that was digitized, this includes an .xml file that reproduces the original digitization in PlotDigitizer. The original manuscript already contained the following statement (though we changed the wording to "used to reproduce" from the original "document"): *For publications, these include the key text passages, tables, maps, or figures that served as the sources for the data. Additional files can be used to reproduce how numeric values were digitized from maps and figures using Plot Digitizer (*.xml) or georeferenced and digitized using QGIS (*.pgw, *.png|jpg.aux.xml, and *.geojson).*